# 第一次作业实验报告

——对数几率回归模型

姓名：汤家军

学号：1611640201

1. 题目

使用python实现对数几率回归模型

1. 问题描述

对鸢尾花数据集实现对数几率回归，实际上是根据已有的数据特征，对数据集进行分类操作，将整个数据集分成0,1两种类型，对其进行拟合操作，验证分类的正确率。

1. 数据集描述

鸢尾花有三个种类，分别是山鸢尾（setosa）、变色鸢尾（versicolor）和维吉尼亚鸢尾（virginica）。该数据集一共包含4个特征变量，1个类别变量。共有150个样本，iris是鸢尾植物，

& Sepal.Length（花萼长度），单位是cm;

& Sepal.Width（花萼宽度），单位是cm;

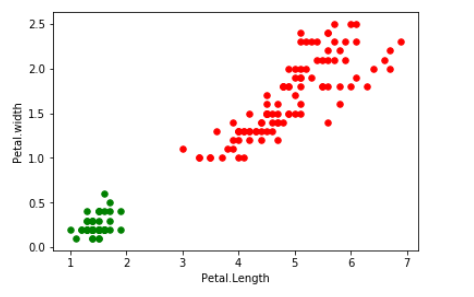
& Petal.Length（花瓣长度），单位是cm;

& Petal.Width（花瓣宽度），单位是cm;

该数据的第一个种类与另外两个种类是线性可分离的，后两个种类是非线性可分离的。

1. 实验结果

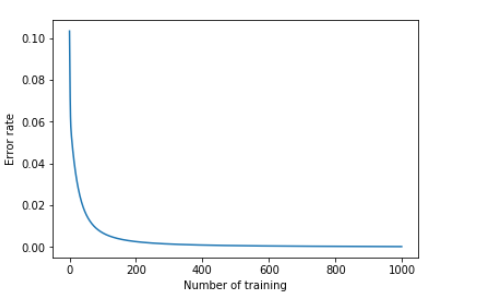
我将鸢尾花数据集分成了两类，第一种种类为1，后两种种类为0，下图是分类后的散点图。



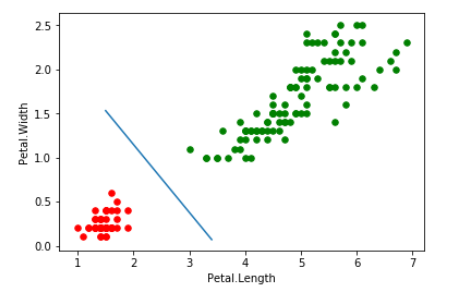
为了使实验后的结果能够更加清晰明了，我取了Sepal length和Sepal width作为特征。

在取数据集的0.5为训练集，0.5为测试集，迭代次数为1000，学习率为0.1时，模型在测试集上的准确率为1.0。

损失函数图像如下：

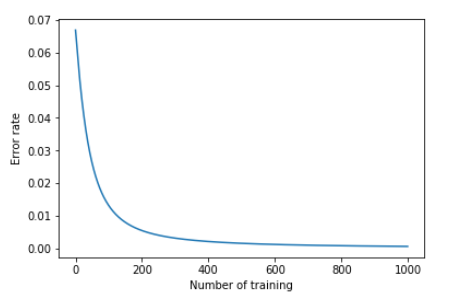


拟合曲线如图：

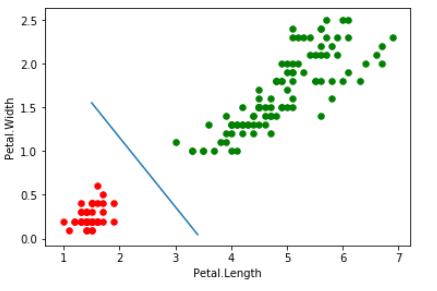


在取数据集的0.7为训练集，0.3为测试集，迭代次数为1000，学习率为0.1时，模型在测试集上的准确率为1.0。

损失函数图像如下：

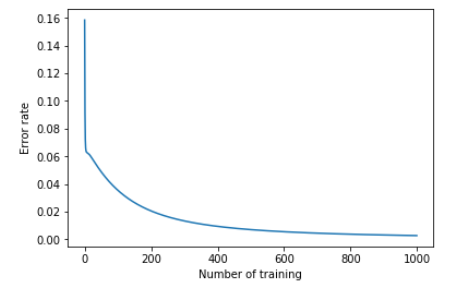


拟合曲线如图：

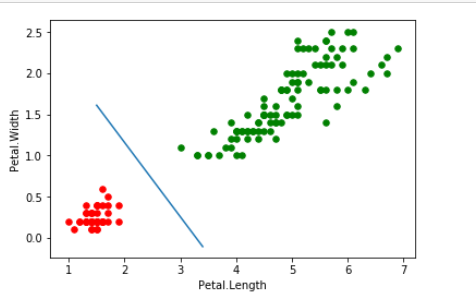


在取数据集的0.9为训练集，0.1为测试集，迭代次数为1000，学习率为0.1时，模型在测试集上的准确率为1.0。

损失函数图像如下：



拟合曲线如图：



1. 实验结果分析

由于数据集的数量太少，导致拟合后的准确率是1.0。

由于给定的theta\_1, theta\_2, theta\_3和学习率alpha的初值，使损失函数的曲线过于陡峭。

由图可以看出，损失函数越来越低，说明随着迭代次数的不断增加，使得准确率越来越高。拟合出来的曲线也将数据集分成了两类。